Sleep Medicine 16 (2015) 173–180

Contents lists available at ScienceDirect

Sleep Medicine

journal homepage: www.elsevier.com/locate/sleep

Original Article

Polysomnography using abbreviated signal montages: impact on sleep and cortical arousal scoring

Warren R. Ruehland [a,b,*], Thomas J. Churchward [a], Linda M. Schachter [c], Tristia Lakey [d], Natalie Tarquinio [e], Fergal J. O'Donoghue [a], Maree Barnes [a], Peter D. Rochford [a]

[a] Institute for Breathing and Sleep, Austin Health, Heidelberg, Vic., Australia
[b] Department of Medicine (Austin Health/Northern Health), University of Melbourne, Heidelberg, Vic., Australia
[c] Sleep Services Australia, Melbourne, Vic., Australia
[d] Advanced Pulmonary and Sleep Diagnostics, Melbourne, Vic., Australia
[e] Bayside Sleep Analysis, Melbourne, Vic., Australia

ARTICLE INFO

Article history:
Received 28 April 2014
Received in revised form 7 October 2014
Accepted 13 November 2014
Available online 26 November 2014

Keywords:
Obstructive sleep apnoea
Diagnosis
Portable monitoring
Accuracy
Precision
Inter-scorer reliability

ABSTRACT

Objective: This study examined the impact of using two abbreviated signal montages on the accuracy, precision and inter-scorer reliability of polysomnography (PSG) sleep and arousal scoring, compared to a standard reference montage, in a cohort of patients investigated for obstructive sleep apnoea (OSA). One abbreviated montage incorporated two signals dedicated to sleep and arousal scoring, and the other incorporated a single signal.
Methods: Four scorers from two laboratories each scored 15 PSGS four times in random order: once using each abbreviated montage and twice using the reference montage.
Results: Use of the two-signal montage resulted in small changes in the distribution of sleep stages, a reduction in the arousal index and resultant reductions in sleep and arousal scoring agreement. For the one-signal montage, although similar magnitude sleep stage distribution changes were observed, there were larger reductions in the arousal index, and sleep and arousal scoring accuracy. Additionally, using the one-signal montage, there were statistically significant reductions in the precision of summary statistics including total sleep time (TST) and the amount of rapid eye movement (REM) sleep scored, and reductions in the inter-scorer reliability of REM sleep and arousal scoring.
Conclusions: These findings demonstrate that abbreviated signal montages may result in underestimation of the arousal index and, depending on the montage, poorer precision in TST and REM sleep scoring, with potential consequences for apnoea–hypopnoea index (AHI) measures and OSA diagnosis. The results highlight the importance of careful evaluation of PSG results when using portable devices that have restricted signals, and they offer guidance for future PSG and portable monitoring standards.

1. Introduction

It is standard clinical practice to confirm obstructive sleep apnoea (OSA) diagnosis using in-laboratory polysomnography (PSG); however, it is increasingly recognised that portable monitoring (PM) may be an acceptable alternative, with acknowledgement and understanding of accompanying limitations [1].

Although many PM devices have abbreviated signal recording capabilities compared to full PSG, there is an advantage in quantifying sleep and cortical arousals, requiring the recording of fast sampling rate signals such as electroencephalography (EEG), electrooculography (EOG), and electromyography (EMG). Recording of these signals allows: (i) assessment of the impact of any respiratory disturbance on sleep architecture; (ii) the use of total sleep time (TST) rather than total recording time (TRT) as the denominator in calculating indices of respiratory or sleep disturbance; (iii) verification of rapid eye movement (REM) sleep sampling, important due to the incidence of REM-related OSA, estimated to have a prevalence of approximately 35% in clinical OSA populations [2]; and (iv) the use of respiratory event scoring criteria requiring airflow reduction accompanied by cortical arousal.

Despite these theoretical advantages, the most recent clinical guidelines for use of PM to diagnose OSA [1] did not consider devices that were capable of measuring sleep. This was because there were no new data available comparing such devices to PSG since previous guidelines [3], stated that evidence was lacking to recommend their clinical use.

* Corresponding author. Institute for Breathing and Sleep, Austin Health, Ground Floor Bowen Centre, Studley Road, Heidelberg, Vic. 3084, Australia. Tel.: +61 3 9496 3528; fax: +61 3 9496 5128.
E-mail address: Warren.Ruehland@austin.org.au (W.R. Ruehland).

http://dx.doi.org/10.1016/j.sleep.2014.11.005
1389-9457/

PM devices with fast sampling capability may still be restricted to the number of signals that can be dedicated to the recording and scoring of sleep and cortical arousals. Current guidelines [4] recommend the use of six primary signals (three EEG, two EOG, and one EMG) for recording and scoring of sleep and cortical arousals in PSG. We have previously shown that the use of four primary signals, incorporating one EEG, results in only small changes in the distribution of sleep stages [5] and no statistically significant differences in sleep or cortical arousal scoring inter-scorer or intra-scorer reliability. However, there are no data to guide clinical practice on how further signal restrictions may impact the scoring of sleep and arousals. Such information is crucial for those using PM devices with limited signal recording capabilities.

This study aimed to examine the impact of using two abbreviated signal montages on the accuracy, precision, and inter-scorer reliability of PSG sleep and arousal scoring, compared to a standard reference montage, in a cohort of patients presenting for the investigation of OSA. One abbreviated montage incorporated two signals dedicated to recording and scoring of sleep and cortical arousals, whereas the other utilised a single signal.

2. Methods

2.1. Design

This study was a prospective, non-blinded, randomised comparison of sleep and arousal scoring using two abbreviated montages compared to a standard reference montage; it was approved by the institutional Human Research Ethics Committee.

2.2. Patient selection

The study utilised 15 single-night PSGs sourced during July and August 2006 from the Austin Health sleep laboratory in Melbourne, Australia, from consecutive patients investigated for OSA. PSGs were not considered if they were primarily conducted for non-OSA sleep disorders, research or treatment implementation.

2.3. PSG recordings

PSGs were recorded using Compumedics S-series or E-series monitoring equipment (Abbotsford, VIC, Australia). The recording configuration consisted of: one EEG signal (C4/A1), two EOG signals (left and right outer canthus (OC)/Fpz), one combined EEG/EOG signal (Fp1/LOC), submental EMG, electrocardiogram (ECG), nasal pressure, body position, thoracic and abdominal excursion (inductance plethysmography), oxygen saturation via finger pulse oximetry (Nellcor N-595; Nellcor Inc, Boulder, CO, USA), left and right leg movement and sound.

2.4. PSG scoring

Sleep and arousal scoring were performed manually, in a single pass, using Profusion PSG 2 software (Compumedics, Abbotsford, VIC, Australia), based on published standards available at the time of the study [6,7]. Apnoea–hypopnoea indices (AHIs) determined using "Chicago Criteria" [8] during the original clinical investigation characterised the patient sample.

During scoring, PSGs were configured to display one of three montages: (i) a reference montage (M_{Ref}) incorporating one EEG signal (C4/A1), two EOG signals and one EMG signal, selected as it was in the minimum configuration recommended for use in PSG [6]; (ii) an abbreviated two-signal montage (M_2) incorporating one EEG signal (C4/A1) and one EOG signal (LOC/Fpz); or (iii) an abbreviated one-signal montage (M_1) incorporating the single combined EEG/EOG signal (Fp1/ROC). Care was taken to ensure that the display size of all signals was identical regardless of the number of signals displayed.

During abbreviated montage scoring, the start and end of REM sleep were not defined by EMG changes, but they were defined by the presence/absence of REMs and stage 2 sleep features; arousals in REM did not require concurrent EMG elevation with EEG frequency shift.

2.5. Scorers

Four scorers from two separate Australian clinical sleep investigation services participated: two from Sleep Services Australia, Melbourne, and two from the Austin Hospital, Melbourne. All scorers participated in scoring concordance programmes and they were experienced in abbreviated montage scoring.

2.6. Protocol

For each scorer, all PSGs and versions were de-identified and presented in random order with the exception that no 2 versions of the same PSG were ever presented consecutively. A second copy of M_{Ref} (M_{Ref2}) was later scored to allow comparison of abbreviated montage accuracy and precision against intra-montage scoring repeatability. Thus, each scorer analysed all 15 PSGs four times each, twice using M_{Ref} and once each using M_1 and M_2.

2.7. Analysis

The analysis involved assessment of: (i) PSG summary statistic accuracy, (ii) PSG summary statistic precision, (iii) event-by-event/epoch-by-epoch accuracy, and (iv) event-by-event/epoch-by-epoch inter-scorer reliability. For all assessments of accuracy and precision, the mean value of all four scorers was used for statistical analysis. Distributions of the differences between numerous parameter pairs were skewed and so non-parametric Friedman tests were undertaken for all comparisons, with post hoc analysis conducted using Wilcoxon signed-rank tests.

2.7.1. Summary statistics accuracy

Statistical analysis compared repeated measure differences in PSG sleep and arousal summary statistics between M_1, M_2, M_{Ref}, and M_{Ref2}. The distribution of epoch-by-epoch sleep stage specific discordances was examined to elucidate the cause of any observed differences.

2.7.2. Summary statistics precision

Precision of PSG sleep and arousal summary statistics for M_1, M_2, and M_{Ref2} each were assessed using the median absolute deviation (MAD) about the median difference from M_{Ref}. Statistical analysis compared repeated measure differences in precision between M_1, M_2, and M_{Ref2}.

2.7.3. Epoch-by-epoch/event-by-event accuracy

Epoch-by-epoch accuracy of sleep and arousal scoring for M_1, M_2, and M_{Ref2} each versus M_{Ref} was assessed using Cohen's pair-wise kappa [9], modified for continuous measurements for arousal scoring [10]. Statistical analysis compared repeated measure differences in accuracy between M_1, M_2, and M_{Ref2}. Raw agreement, expressed as percentage agreement [11] for sleep and as proportion of specific agreement (PSA) for positive ratings [12] for arousals, was also presented for comparison.

2.7.4. Epoch-by-epoch/event-by-event inter-scorer reliability

Epoch-by-epoch inter-scorer reliability of sleep and arousal scoring for M_1, M_2, M_{Ref}, and M_{Ref2} each were assessed using Fleiss' multi-scorer kappa [11,12], modified for continuous measurements for

Table 1
Patient characteristics.

Parameter	Value	IQR	Range
N	15	–	–
Gender (M/F)	10/5	–	–
Age (years)	51	39, 56	33, 71
BMI (kg/m [2])	32.4	30.1, 36.5	23.7, 54.0
ESS	9	7, 12	5, 15
AHI (/h)	12.6	10.5, 24.7	1.6, 80.6

Values are median where appropriate.
Abbreviations: IQR inter-quartile range; BMI = body mass index; ESS = Epworth Sleepiness Scale; AHI = apnoea–hypopnoea Index.

arousal scoring. Statistical analysis compared repeated measure differences in reliability between M_1, M_2, M_{Ref}, and M_{Ref2}. Raw agreement was also presented for comparison as described above.

All data are represented as median (inter-quartile range). A *P* value of <0.05 was accepted as statistically significant.

3. Results

3.1. Patient characteristics

The characteristics of the 15 patients studied are presented in Table 1.

3.2. Summary statistics accuracy

3.2.1. M_2 *versus* M_{Ref}

There was a statistically significant reduction of 29 (20, 39)% in stage 1 sleep and a statistically significant increase of 25 (9, 33)% in slow-wave sleep (SWS); additionally, there was a statistically significant arousal index reduction of 18 (6, 21)% (Table 2; Fig. 1). No statistically significant difference was found for other sleep summary statistics such as TST (Fig. 2). The reduction in stage 1 sleep was largely due to a net shift from stage 1 to stage 2 of 14.5 (9.4, 18) min. The increase in SWS was largely due to a net shift from stage 2 to SWS of 9.5 (3.0, 23.4) min.

Table 2
Accuracy of polysomnography sleep and arousal scoring summary statistics derived using two abbreviated signal montages (M_1 and M_2) compared to scoring using a full montage (M_{Ref}). Accuracy of repeated scoring of the full montage (M_{Ref2}) is also reported for comparison.

Parameter	Montage	Value	Difference Relative to M_{Ref}	*P*-Value
Total sleep time (min)	M_{Ref}	302.9 (247.9, 340.9)	–	
	M_{Ref2}	306.9 (248.4, 354.3)	3.1 (–1.1, 6.3)	NS
	M_2	308.3 (245.9, 346.1)	3.1 (1.8, 4.9)	NS
	M_1	318.1 (236.4, 345.7)	–3.3 (–7.8, 15.8)	NS
Sleep efficiency (%)	M_{Ref}	71.8 (61.4, 77.7)	–	–
	M_{Ref2}	72.9 (61.6, 79.3)	0.6 (–0.3, 1.4)	NS
	M_2	72.2 (64.3, 79.1)	0.8 (0.5, 1.2)	NS
	M_1	76.5 (62.4, 80.4)	–0.6 (–2.5, 3.6)	NS
Sleep latency (min)	M_{Ref}	19.3 (13.3, 37.8)	–	–
	M_{Ref2}	21.6 (12.8, 36.3)	0.0 (–0.6, 2.9)	NS
	M_2	19.1 (13.2, 34.9)	–0.1 (–0.4, 0.1)	NS
	M_1	23.8 (16.3, 37.0)	1.9 (0.6, 3.8)	NS
REM latency (min)	M_{Ref}	95.6 (63.9, 152.9)	–	–
	M_{Ref2}	96.1 (68.1, 145.7)	0.5 (–0.4, 1.4)	0.198
	M_2	98.3 (70.2, 161.7)	1.9 (0.1, 5.9)	0.100
	M_1	87.6 (69.8, 107.3)	–2.0 (–16.1, –0.3)	0.069
Wake after sleep onset (min)	M_{Ref}	94.8 (83.2, 118.1)	–	–
	M_{Ref2}	93.8 (80.9, 109.9)	–3.8 (–5.9, 1.1)	NS
	M_2	97.0 (79.8, 114.9)	–3.1 (–4.4, –1.4)	NS
	M_1	87.4 (70.0, 114.8)	0.1 (–13.7, 5.4)	NS
Time in each sleep stage (min)				
Stage 1	M_{Ref}	23.6 (20.0, 35.5)	–	–
	M_{Ref2}	22.9 (20.8, 35.5)	1.8 (–1.4, 5.1)	0.147
	M_2	15.5 (12.3, 29.4)	–7.3 (–9.7, –5.3)	**0.001**
	M_1	13.0 (10.8, 25.8)	–8.8 (–12.1, –4.8)	**0.002**
Stage 2	M_{Ref}	167.3 (137.7, 189.3)	–	–
	M_{Ref2}	168.0 (127.2, 182.6)	–6.1 (–14.6, 0.1)	0.078
	M_2	176.8 (126.8, 199.6)	7.3 (–3.7, 9.3)	0.164
	M_1	174.4 (145.2, 226.0)	10.3 (1.6, 24.9)	**0.020**
Slow wave sleep	M_{Ref}	45.3 (25.6, 52.8)	–	–
	M_{Ref2}	47.3 (23.9, 68.0)	4.5 (–0.5, 10.4)	**0.044**
	M_2	54.0 (25.3, 64.3)	6.3 (3.3, 11.8)	**0.006**
	M_1	32.6 (14.9, 43.4)	–7.6 (–18.1, 3.4)	0.211
REM	M_{Ref}	51.0 (40.8, 68.4)	–	–
	M_{Ref2}	51.0 (41.4, 69.3)	0.0 (–0.3, 1.5)	0.220
	M_2	49.4 (36.6, 68.8)	–1.4 (–2.6, –0.0)	0.057
	M_1	52.1 (36.6, 80.5)	3.6 (1.1, 12.3)	0.078
Arousal index (/h)	M_{Ref}	15.2 (13.2, 19.2)	–	–
	M_{Ref2}	16.3 (14.1, 18.6)	–0.2 (–1.5, 2.0)	0.955
	M_2	12.9 (10.8, 16.9)	–2.7 (–4.9, –0.7)	**0.002**
	M_1	9.7 (5.6, 14.1)	–6.0 (–8.6, –4.2)	**0.001**
Arousal Count	M_{Ref}	77.2 (54.9, 100.9)	–	–
	M_{Ref2}	83.0 (64.1, 88.3)	3.4 (–3.7, 11.1)	0.363
	M_2	66.3 (50.1, 87.9)	–11.7 (–22.0, –3.2)	**0.006**
	M_1	42.2 (29.5, 70.2)	–28.0 (–37.2, –18.2)	**0.001**

Data reported as median (inter-quartile range); abbreviations: M_{Ref}: reference montage; M_2: one central EEG, one EOG montage; M_1: one combined EEG/EOG montage; M_{Ref2}: repeated scoring of M_{Ref}; NS: $P > 0.05$ using Friedman test.

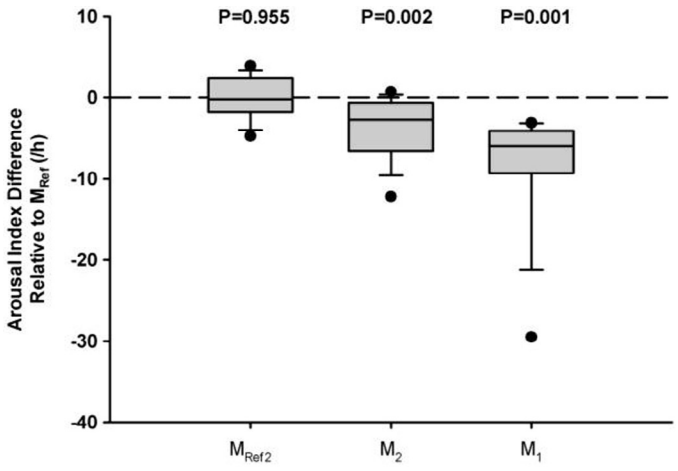

Fig. 1. Box plots showing paired arousal index differences when arousal scoring using M_{Ref2}, M_2 and M_1 versus M_{Ref} ($n = 15$ PSGs). Abbreviations: M_{Ref}: reference montage; M_2: one central EEG, one EOG montage; M_1: one combined EEG/EOG montage; M_{Ref2}: repeated scoring of M_{Ref}. P value determined using Wilcoxon signed-rank test following statistically significant Friedman test. Central line = median; box boundary = 25th and 75th percentiles; error bars = 10th and 90th percentiles; ● = outliers.

3.2.2. M_1 *versus* M_{Ref}

There was a statistically significant reduction of 39 (21, 50)% in stage 1 sleep and a statistically significant increase of 8 (0, 15)% in stage 2 sleep; there was also a statistically significant arousal index reduction of 41 (28, 51)% (Table 2; Fig. 1). No statistically significant difference was found for other sleep summary statistics. For the arousal index, the median reductions were greater for M_1 compared to M_2 versus M_{Ref} ($P = 0.001$). A net shift from stage 1 to stage 2 of 11.3 (3.0, 16.9) min and a net shift from stage 1 to awake of 7.8 (–0.9, 10.8) min contributed to the reduction in stage 1. The shift from stage 1 to stage 2 as well as a net shift from SWS to stage 2 sleep of 14.8 (–1.9, 36.8) min contributed to the increase in stage 2.

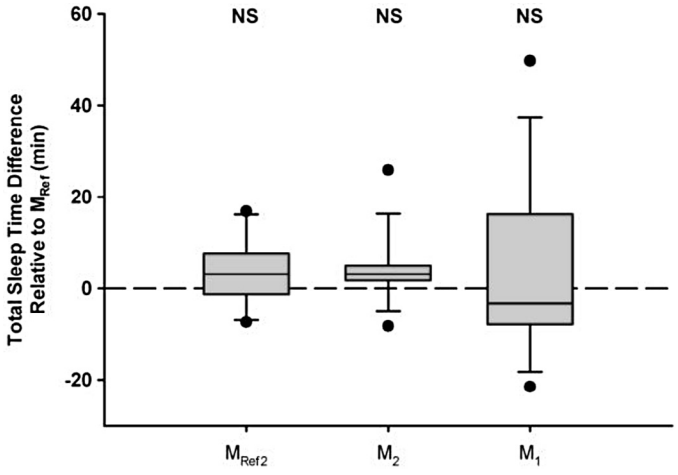

Fig. 2. Box plots showing paired total sleep time differences when sleep scoring using M_{Ref2}, M_2 and M_1 versus M_{Ref} ($n = 15$ PSGs). Abbreviations: M_{Ref}: reference montage; M_2: one central EEG, one EOG montage; M_1: one combined EEG/EOG montage; M_{Ref2}: repeated scoring of M_{Ref}; NS: $P > 0.05$ using Friedman test. Central line = median; box boundary = 25th and 75th percentiles; error bars = 10th and 90th percentiles; ● = outliers.

3.2.3. M_{Ref2} *versus* M_{Ref}

There was a statistically significant increase of 14 (–1, 34)% in SWS, largely due to a net shift from stage 2 to SWS of 7.8 (–2.6, 17.4) min.

3.3. *Summary statistics precision*

3.3.1. M_2 *versus* M_{Ref}

There were no statistically significant differences in precision compared to repeated scoring of the full montage (M_{Ref2} vs. M_{Ref}), for any sleep or arousal summary statistics (Table 3).

3.3.2. M_1 *versus* M_{Ref}

There was a statistically significant reduction in precision (ie, an increase in MAD) for TST, SE, RL, and the amount of stage REM sleep, compared to the precision of M_{Ref2} (Table 3). The reduced precision of M_1 for those parameters is also reflected in the wider interquartile ranges of the differences from M_{Ref} (Table 2; Fig. 2).

3.4. *Epoch-by-epoch/event-by-event accuracy*

There was a statistically significant reduction in the accuracy of M_2 versus M_{Ref} compared to M_{Ref2} versus M_{Ref}, for overall sleep scoring, and for individual sleep stages: stage 1, stage 2, and REM sleep (Table 4). There was a statistically significant reduction in the accuracy of M_1 versus M_{Ref} compared to M_{Ref2} versus M_{Ref} for overall sleep scoring, and for all sleep stages individually. For overall sleep scoring, the reduction in accuracy was greater for M_1 compared to M_2 ($P = 0.001$), and the equivalent raw percentage agreements were 86.0% (82.7, 86.8), 79.7% (75.0, 83.2) and 86.6% (84.5, 89.0) using M_2, M_1 and M_{Ref2}, respectively. The arousal scoring accuracy was reduced for both M_1 and M_2 versus M_{Ref} compared to M_{Ref2} versus M_{Ref} (Table 4), and the reduction was greater for M_1 compared to M_2 ($P = 0.001$). For arousal scoring raw agreement, the equivalent PSA were 0.67 (0.64, 0.75), 0.55 (0.47, 0.61) and 0.73 (0.70, 0.78) when using M_2, M_1 and M_{Ref2}, respectively.

3.5. *Inter-scorer reliability*

There were no statistically significant reductions in multi-rater inter-scorer reliability of sleep or arousal scoring when scoring with either M_2 or M_{Ref2} versus M_{Ref} (Table 5). With M_1 versus M_{Ref}, there was a statistically significant inter-scorer reliability reduction for scoring REM sleep and arousals. For overall sleep scoring, the equivalent raw percentage agreements were 79.8% (77.0, 83.8), 78.7% (75.3, 0.81), 80.2% (77.3, 84.1) and 79.2% (75.6, 82.1) using M_2, M_1, M_{Ref2}, and M_{Ref}, respectively. For arousal scoring raw agreement, the equivalent PSA were 0.58 (0.56, 0.66), 0.60 (0.49, 0.62), 0.63 (0.58, 0.73) and 0.66 (0.58, 0.72) when using M_2, M_1, M_{Ref2} and M_{Ref}, respectively.

4. Discussion

This study examined the impact of using two abbreviated signal montages on the accuracy, precision, and inter-scorer reliability of sleep and arousal scoring, compared to a standard reference montage. Using an abbreviated montage incorporating one central EEG signal and one EOG signal for sleep and arousal scoring, we found that there was an increase in the amount of SWS and a reduction in stage 1 sleep, as well as a reduction in the cortical arousal index. Despite a small but statistically significant reduction in epoch-by-epoch sleep scoring and event-by-event arousal scoring accuracy, there was no statistically significant difference in the precision of PSG summary statistics compared to repeat scoring using the reference montage. Additionally, there were no statistically significant differences in multi-rater inter-scorer reliability of sleep or arousal scoring compared to the reference montage.

Table 3
Precision of sleep and arousal scoring summary statistics derived using two abbreviated signal montages (M_1 and M_2 vs. M_{Ref}) compared to repeated scoring using a full montage (M_{Ref2} vs. M_{Ref}).

Parameter	Montage	Value	Difference Relative to M_{Ref2}	*P*-Value
Total sleep time (min)	M_{Ref2}	4.4 (1.4, 7.8)	–	–
	M_2	1.6 (1.2, 5.8)	–0.9 (–4.1, 1.6)	0.394
	M_1	10.3 (3.7, 19.0)	7.4 (–1.3, 13.8)	**0.035**
Sleep efficiency (%)	M_{Ref2}	0.9 (0.5, 2.2)	–	–
	M_2	0.3 (0.3, 1.5)	–0.5 (–1.2, 0.5)	0.233
	M_1	2.4 (1.1, 4.5)	1.6 (0.0, 3.1)	**0.029**
Sleep latency (min)	M_{Ref2}	0.9 (0.5, 4.9)	–	–
	M_2	0.3 (0.2, 1.1)	–0.3 (–4.5, 0.3)	NS
	M_1	1.6 (0.8, 3.0)	–0.3 (–2.6, 2.6)	NS
Stage REM latency (min)	M_{Ref2}	0.9 (0.4, 2.6)	–	–
	M_2	2.1 (1.1, 16.3)	0.9 (–0.3, 2.1)	0.158
	M_1	9.9 (1.6, 18.1)	3.6 (0.8, 13.0)	**0.005**
Wake after sleep onset (min)	M_{Ref2}	4.3 (1.5, 8.1)	–	–
	M_2	1.6 (1.1, 4.8)	–1.0 (–4.6, 0.5)	NS
	M_1	8.6 (3.3, 17.3)	4.9 (–3.9, 10.7)	NS
Time in each sleep stage (min)				
Stage 1	M_{Ref2}	3.3 (1.4, 5.2)	–	–
	M_2	2.4 (1.4, 4.1)	–1.5 (–3.1, 2.6)	NS
	M_1	3.8 (2.8, 5.9)	0.5 (–1.9, 4.8)	NS
Stage 2	M_{Ref2}	7.8 (4.4, 11.9)	–	–
	M_2	8.6 (1.6, 16.4)	–0.4 (–6.4, 7.8)	NS
	M_1	11.8 (4.8, 19.7)	1.6 (–5.5, 11.3)	NS
Slow-wave sleep	M_{Ref2}	6.9 (1.5, 11.0)	–	–
	M_2	5.5 (2.8, 12.3)	–0.8 (–5.8, 2.4)	NS
	M_1	11.8 (5.7, 17.2)	2.5 (–5.0, 9.8)	NS
REM	M_{Ref2}	0.5 (0.3, 1.5)	–	–
	M_2	1.4 (0.8, 2.8)	0.9 (–0.3, 2.3)	0.078
	M_1	8.0 (1.8, 9.3)	3.6 (1.0, 9.1)	**0.003**
Arousal index (/h)	M_{Ref2}	1.7 (0.9, 3.0)	–	–
	M_2	2.1 (0.6, 3.7)	0.0 (–0.6, 1.4)	NS
	M_1	2.0 (1.0, 3.1)	0.4 (–0.5, 1.7)	NS
Arousal Count	M_{Ref2}	7.3 (4.8, 10.2)	–	–
	M_2	9.2 (6.5, 16.9)	1.1 (–0.5, 6.0)	NS
	M_1	9.3 (6.0, 14.4)	5.1 (–2.6, 8.6)	NS

All data reported as median (inter-quartile range) absolute deviation around the median difference from M_{Ref}; abbreviations: M_{Ref}: reference montage; M_2: one central EEG, one EOG montage; M_1: one combined EEG/EOG montage; M_{Ref2}: repeated scoring of M_{Ref}; NS: $P > 0.05$ using Friedman test.

Table 4
Epoch-by-epoch and event-by-event accuracy (kappa) of polysomnography sleep and arousal scoring using two abbreviated signal montages relative to an a full montage (M_1 and M_2 vs. M_{Ref}) in comparison to repeat scoring of a full montage (M_{Ref2} vs. M_{Ref}).

Parameter	Montage Comparison	Value	Difference Relative to M_{Ref2}	*P*-Value
Sleep – Overall				
Awake/1/2/SWS/REM	M_{Ref2} versus M_{Ref}	0.80 (0.78, 0.82)	–	–
	M_2 versus M_{Ref}	0.79 (0.75, 0.81)	–0.03 (–0.04, –0.01)	**0.004**
	M_1 versus M_{Ref}	0.69 (0.64, 0.75)	–0.11 (–0.13, –0.07)	**0.001**
Sleep – Specific Sleep Stages				
Stage 1	M_{Ref2} versus M_{Ref}	0.49 (0.47, 0.56)	–	–
	M_2 versus M_{Ref}	0.42 (0.38, 0.46)	–0.08 (–0.12, –0.05)	**0.002**
	M_1 versus M_{Ref}	0.27 (0.23, 0.34)	–0.21 (–0.24, –0.19)	**0.001**
Stage 2	M_{Ref2} versus M_{Ref}	0.79 (0.73, 0.81)	–	–
	M_2 versus M_{Ref}	0.75 (0.70, 0.79)	–0.04 (–0.05, –0.01)	**0.009**
	M_1 versus M_{Ref}	0.67 (0.58, 0.77)	–0.10 (–0.16, –0.06)	**0.001**
Slow-wave sleep	M_{Ref2} versus M_{Ref}	0.72 (0.53, 0.79)	–	–
	M_2 versus M_{Ref}	0.68 (0.52, 0.75)	–0.02 (–0.05, 0.01)	0.211
	M_1 versus M_{Ref}	0.51 (0.45, 0.71)	–0.13 (–0.20, –0.06)	**0.001**
REM	M_{Ref2} versus M_{Ref}	0.92 (0.89, 0.94)	–	
	M_2 versus M_{Ref}	0.89 (0.80, 0.91)	–0.03 (–0.06, –0.02)	**0.001**
	M_1 versus M_{Ref}	0.71 (0.67, 0.85)	–0.14 (–0.29, –0.08)	**0.001**
Awake	M_{Ref2} versus M_{Ref}	0.88 (0.84, 0.92)	–	–
	M_2 versus M_{Ref}	0.89 (0.85, 0.91)	0.00 (–0.01, 0.01)	0.609
	M_1 versus M_{Ref}	0.80 (0.71, 0.83)	–0.09 (–0.12, –0.06)	**0.001**
Arousals – Overall	M_{Ref2} versus M_{Ref}	0.66 (0.61, 0.69)	–	–
	M_2 versus M_{Ref}	0.59 (0.55, 0.67)	–0.04 (–0.07, –0.02)	**0.001**
	M_1 versus M_{Ref}	0.46 (0.41, 0.51)	–0.19 (–0.24, –0.15)	**0.001**

All data reported as median (inter-quartile range); abbreviations: M_{Ref}: reference montage; M_2: one central EEG, one EOG montage; M_1: one combined EEG/EOG montage; M_{Ref2}: repeated scoring of M_{Ref}.

Table 5
Epoch-by-epoch and event-by-event inter-rater reliability (kappa) of polysomnography sleep and arousal scoring derived using two abbreviated signal montages (M_1 and M_2) compared to scoring using a full montage (M_{Ref}). Inter-rater reliability of the second scoring of the full montage (M_{Ref2}) is also reported for comparison.

Parameter	Montage	Value	Difference Relative to M_{Ref}	*P*-Value
Sleep – Overall				
Awake/1/2/SWS/REM	M_{Ref}	0.71 (0.65, 0.74)	–	–
	M_{Ref2}	0.72 (0.68, 0.77)	0.01 (0.00, 0.03)	NS
	M_2	0.70 (0.67, 0.77)	0.01 (0.00, 0.03)	NS
	M_1	0.68 (0.64, 0.71)	–0.01 (–0.06, 0.04)	NS
Sleep – Specific sleep stages				
Stage 1	M_{Ref}	0.31 (0.26, 0.36)	–	–
	M_{Ref2}	0.32 (0.25, 0.35)	–0.02 (–0.03, 0.02)	NS
	M_2	0.25 (0.20, 0.32)	–0.06 (–0.08, –0.01)	NS
	M_1	0.26 (0.18, 0.32)	–0.07 (–0.13, 0.02)	NS
Stage 2	M_{Ref}	0.66 (0.54, 0.72)	–	
	M_{Ref2}	0.68 (0.60, 0.74)	0.02 (–0.01, 0.06)	NS
	M_2	0.65 (0.59, 0.73)	0.02 (0.00, 0.07)	NS
	M_1	0.64 (0.60, 0.71)	0.00 (–0.05, 0.08)	NS
Slow-wave sleep	M_{Ref}	0.57 (0.33, 0.68)	–	–
	M_{Ref2}	0.63 (0.46, 0.80)	0.05 (–0.02, 0.22)	0.053
	M_2	0.64 (0.53, 0.73)	0.07 (0.00, 0.14)	**0.041**
	M_1	0.50 (0.25, 0.59)	–0.13 (–0.16, 0.00)	0.078
REM	M_{Ref}	0.87 (0.80, 0.91)	–	
	M_{Ref2}	0.90 (0.82, 0.91)	0.02 (0.00, 0.03)	0.100
	M_2	0.84 (0.68, 0.92)	–0.03 (–0.08, 0.00)	0.090
	M_1	0.74 (0.57, 0.82)	–0.11 (–0.22, –0.04)	**0.030**
Awake	M_{Ref}	0.85 (0.77, 0.89)	–	–
	M_{Ref2}	0.84 (0.78, 0.89)	0.00 (–0.02, 0.01)	NS
	M_2	0.85 (0.80, 0.89)	0.00 (–0.02, 0.02)	NS
	M_1	0.81 (0.72, 0.84)	–0.04 (–0.08, 0.03)	NS
Arousals – Overall	M_{Ref}	0.53 (0.46, 0.57)	–	–
	M_{Ref2}	0.54 (0.45, 0.58)	0.01 (0.00, 0.03)	0.382
	M_2	0.48 (0.46, 0.56)	–0.02 (–0.05, 0.01)	0.191
	M_1	0.46 (0.41, 0.49)	–0.05 (–0.13, –0.02)	**0.018**

All data reported as median (inter-quartile range); abbreviations: M_{Ref}: reference montage; M_2: one central EEG, one EOG montage; M_1: one combined EEG/EOG montage; M_{Ref2}: repeated scoring of M_{Ref}; NS: $P > 0.05$ using Friedman test.

Similar to the two-signal montage, using the one-signal montage resulted in a change in the distribution of sleep stages: in particular, a reduction in stage 1 sleep, and an increase in stage 2 sleep. However, compared to the two-signal montage, there was a greater reduction in arousals scored and the reductions in epoch-by-epoch and event-by-event accuracy were larger and involved more sleep stages. Additionally, using the one-signal montage, there were reductions in the precision of sleep scoring summary statistics (TST, SE, REM latency, and total REM), and there were reductions in inter-scorer reliability of REM sleep and arousal scoring.

4.1. Sleep summary statistics

It is of interest that the two-signal montage resulted in an increase in SWS sleep whereas the single-signal montage did not. One may expect that, given that electrodes are placed more frontally in the single-signal montage and that slow waves used to score SWS are more prominent in frontal regions of the cortex [13], there would be an increase in scored SWS. Our finding is possibly contributed by the electrode being placed too far forward to pick up frontal lobe predominant slow waves, or it is due to the smaller inter-electrode distance, which is known to result in EEG amplitude reduction [14]. For the two-signal montage increase in SWS, given that reference montage rescoring also resulted in a small but significant increase in SWS, one must also consider the possibility of a false-positive finding. While it seems plausible that eye movement-related EEG fluctuations being mistaken for K-complexes may contribute to the reductions in stage 1 sleep for both abbreviated montages, if this were true, one might also expect an increase in stage 2 at the expense of REM sleep, which was not evident in the present study. Rather, the reduction in precision for both abbreviated montages in the amount of REM sleep scored, significantly so for the one-signal montage, suggests confusion between eye movement deflections and K-complexes but not one being consistently mistaken for the other.

A bias or shift in the distribution of sleep stages may not be a great concern but a decrease in precision is more worrisome as the impact of using an abbreviated montage becomes less predictable or systematic. Precision was only reduced using the one-signal montage, significantly so for TST, sleep efficiency, REM latency, and the amount of REM sleep scored. The reduced precision for TST is of particular concern, given its use as the denominator in the AHI and ArI.

4.2. Sleep epoch-by-epoch accuracy

A reduction in epoch-by-epoch accuracy using abbreviated montages is expected given the shifts in sleep stages observed for both abbreviated montages and the reduced precision observed for the one-signal montage. The median reduction in overall epoch-by-epoch sleep scoring accuracy for the two-signal montage was similar in magnitude to the small average difference in inter-scorer reliability seen when sleep scoring with one versus three EEGs [5] or when using Rechtschaffen and Kales (R&K) [6] versus American Academy of Sleep Medicine (AASM) [4] sleep scoring recommendations [15]. The median reduction for the one-signal montage was three to four times greater than the two-signal montage. Additionally, individual sleep stage epoch-by-epoch accuracy reductions were greater when using the one-signal montage compared to the two-signal montage.

4.3. Sleep inter-scorer reliability

The only statistically significant reduction in sleep inter-scorer reliability was for REM sleep using the one-signal montage.

Together with the lack of precision in the amount of REM scored and the reduced REM epoch-by-epoch accuracy, this finding when using the one-signal montage opposes any advantage gained in being able to identify REM sleep.

4.4. Cortical arousal summary statistics

It is worth noting that both abbreviated montages resulted in reductions in the number of scored arousals. Given that the two-signal montage used the same EEG electrode placement as the reference montage, the reduction observed likely relates to the lack of EMG. Even though EMG activation is only a requirement for arousal scoring in REM sleep, as previously suggested [16], EMG activation in NREM sleep may act to cue for closer inspection of EEG. The further reduction in scored arousals using the one-signal montage likely then relates to electrode placement. This further reduction is inconsistent with a previous finding suggesting an increase in arousals with frontal leads [17], perhaps indicating that the combined frontal EEG/EOG is too far forward to effectively detect all frontal lobe arousals. Although not examined in this study, due to the reduction in arousals scored, a reduction in AHI is likely when hypopnoea scoring is dependent on cortical arousal association. The magnitude of this AHI reduction requires further investigation; however, we were able to examine data from a previous study [18] where we determined that a median of 35% (12, 50) of hypopnoeas were scored based on the presence of an associated arousal alone when using the 2007 AASM alternative hypopnoea definition [4]. Applying the median reduction of 16.2% in the arousal count observed in the present study using the two-signal montage resulted in a reduction in an AHI of 0.8 (0.2, 1.6)/h or 4.7 (1.1, 7.2)%; applying the median arousal count reduction of 42.6% observed in the present study using the one-signal montage resulted in an AHI reduction of 2.1 (0.4, 4.1)/h or 12.3 (2.9, 19.0)%. Care must be exercised when making such a comparison, however, given the possible differences in the sampled populations. Indeed, by chance, the subjects selected in the current study had a relatively low median AHI of 12.6/h compared to 25.1/h in our previous study [18]. This observation also highlights the need for care when extrapolating the results of the present study to more severe OSA populations.

4.5. Cortical arousal event-by-event accuracy

The present finding of a reduction in event-by-event arousal scoring accuracy for both abbreviated montages is expected given the lower number of arousals scored.

4.6. Cortical arousal inter-scorer reliability

Both abbreviated montages resulted in lower inter-scorer reliability, but only significantly so for the one-signal montage. This reduction in inter-scorer reliability is a concern given that inter-scorer reliability of arousal scoring is already generally poorer than sleep or respiratory event scoring [5]. Again, this finding may at least in part be explained by the lack of EMG cue in the abbreviated montages; however, this does not explain the larger reliability decline using the one-signal abbreviated montage. Reduced inter-scorer reliability of arousal scoring also has implications for the reliability of scoring cortical arousal-associated respiratory events.

4.7. Study limitations

One potential limitation of the current study is that it was commenced prior to the publication of current standards [4], and therefore the study used recording and scoring techniques based on previous guidelines [6,7]. However, as noted by the Italian Association of Sleep Medicine [19], the AASM manual's sleep and arousal scoring specifications retain much of the framework of previous guidelines. Thus, only small changes in sleep stage distributions have been observed when comparing AASM to R&K scoring [20], or multiple to single EEG derivations [5], suggesting that the results obtained in the present study are applicable to current guidelines.

A limitation that must also be recognised in the present study is that scorers could not be blinded to scoring method and therefore could be subject to bias. We tried to limit this by using multiple scorers from different sleep centres, which also had the advantage of increasing the generalisability of the study findings.

One advantage of the one-signal montage used in the present study is that the electrodes sit outside the hairline, which allows the use of pre-gelled adhesive electrodes, simplifying electrode application and removal. This advantage comes at the cost of reduced accuracy, precision and reliability of sleep scoring, and reduced accuracy and reliability of cortical arousal scoring; however, it is important to recognise that this does not necessarily preclude the use of single-electrode montages with different electrode placement. For example, Dyson et al. [21] compared sleep scoring with EEG electrode placement outside the hairline (~Fp2/A1) to standard R&K placement (C4/A1) and reported comparable sleep scoring. Thus, further studies are required investigating the utility of different electrode placements to those used in the present study. Additionally, it remains unknown whether non-epoch-based scoring approaches such as cyclical alternating pattern (CAP) [22], which better account for sleep microstructure such as K-complexes, sleep spindles and delta bursts, would also be impacted by abbreviated montages. One may speculate that information would be lost by limiting the channels available for analysis.

It is also important to recognise that non-statistically significant findings in the present study do not necessarily imply the equivalence of methods; it is possible that with a larger sample significant differences would be found. It is clear, however, that more statistically significant differences were observed for the one-signal montage compared to the two-signal montage, showing that differences were larger and/or more consistent.

5. Conclusion

The main findings of this study were that using an abbreviated montage of one central EEG and one EOG during PSG in a cohort of patients being investigated for OSA resulted in a change in the distribution in sleep stages and a reduction in the arousal index and thus a small reduction in epoch-by-epoch and event-by-event accuracy. Similar to the two-signal montage, the one-signal montage resulted in a change in the distribution of sleep stages; however, compared to the two-signal montage, using the one-signal montage, the reduction in the arousal index was greater, and the reductions in epoch-by-epoch and event-by-event accuracy were larger and involved more sleep stages. Additionally, using the one-signal montage, there were reductions in the precision of sleep scoring summary statistics and inter-scorer reliability of sleep and arousal scoring, not observed using the two-signal montage. The reduction in precision of TST is of particular concern given its use as the denominator of AHI and AI.

These findings are valuable for those using portable devices that have restricted signal options, and they offer guidance for future standards for recording and scoring sleep and related events. They reflect greater uncertainty in scoring sleep and arousals using one compared to two signals during PSG and bring into question the utility in scoring sleep and arousals using a single signal, although it should be recognised that alternative electrode placement or subject populations may alter these findings.

Disclosure statement

Mr Ruehland has received research support from Resmed, Fisher and Paykel Healthcare and Philips Respironics. Mr Ruehland and Mr Rochford are directors of Respiratory QA Pty Ltd (trading as QSleep), which provides quality control services for respiratory and sleep laboratories. Dr O'Donoghue is a partner in a private sleep laboratory service as well as a partner in a business that provides home polysomnography. The remaining authors have declared no conflict of interest.

Conflict of interest

The ICMJE Uniform Disclosure Form for Potential Conflicts of Interest associated with this article can be viewed by clicking on the following link: http://dx.doi.org/10.1016/j.sleep.2014.11.005.

Acknowledgements

Financial Support: National Health and Medical Research Council of Australia (grant numbers 430300 and 430302).

References

[1] Collop N, Anderson W, Boehlecke B, Claman D, Goldberg R, Gottlieb D, et al. Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients. J Clin Sleep Med 2007;3:737–47.
[2] Haba-Rubio J, Janssens J, Rochat T, Sforza E. Rapid eye movement-related disordered breathing – Clinical and polysomnographic features. Chest 2005;128:3350.
[3] Kushida CA, Littner MR, Morgenthaler T, Alessi CA, Bailey D, Coleman J Jr, et al. Practice parameters for the indications for polysomnography and related procedures: an update for 2005. Sleep 2005;28:499–521.
[4] Iber C, Ancoli-Israel S, Chesson A, Quan S, for the American Academy of Sleep Medicine. The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications. 1st ed. Westchester, Ill: American Academy of Sleep Medicine; 2007.
[5] Ruehland W, O'Donoghue F, Pierce R, Thornton A, Singh P, Copland J, et al. The 2007 AASM Recommendations for EEG Electrode Placement in Polysomnography: impact on Sleep and Cortical Arousal Scoring. Sleep 2011;34:73–81.
[6] Rechtschaffen A, Kales A. A manual of standardised terminology, techniques and scoring system for sleep stages of human subjects. Los Angeles: Brain Information Service, Brain Research Institute, UCLA; 1968.
[7] Bonnet MH, Carley D, Carskadon MA, Easton P, Guilleminault C, Harper R, et al. EEG arousals: scoring rules and examples: a preliminary report from the Sleep Disorders Atlas Task Force of the American Sleep Disorders Association. Sleep 1992;15:173–84.
[8] Flemons WW, Buysse D, Redline S, Pack A, Strohl K, Wheatley J, et al. Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. The report of an american academy of sleep medicine task force. Sleep 1999;22:667–89.
[9] Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20:37–46.
[10] Conger AJ. Kappa reliabilities for continuous behaviors and events. Educ Psychol Meas 1985;45:861–8.
[11] Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull 1971;76:378–82.
[12] Fleiss JL. The measurement of Interrater Agreement. Statistical methods for rates and proportions. 2nd ed. New York: John Wiley & Sons; 1981. p. 212–36.
[13] Happe S, Anderer P, Gruber G, Klosch G, Saletu B, Zeitlhofer J. Scalp topography of the spontaneous K-complex and of delta-waves in human sleep. Brain Topogr 2002;15:43–9.
[14] Epstein CM, Brickley GP. Interelectrode distance and amplitude of the scalp EEG. Electroencephalogr Clin Neurophysiol 1985;60:287–92.
[15] Danker-Hopfe H, Anderer P, Zeitlhofer J, Boeck M, Dorn H, Gruber G, et al. Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard. J Sleep Res 2009;18:74–84.
[16] Loredo JS, Clausen JL, Ancoli-Israel S, Dimsdale JE. Night-to-night arousal variability and interscorer reliability of arousal measurements. Sleep 1999;22:916–20.
[17] O'Malley EB, Norman RG, Farkas D, Rapoport DM, Walsleben JA. The addition of frontal EEG leads improves detection of cortical arousal following obstructive respiratory events. Sleep 2003;26:435–9.
[18] Ruehland WR, Rochford PD, O'Donoghue FJ, Pierce RJ, Singh P, Thornton AT. The new AASM criteria for scoring hypopneas: Impact on the apnea hypopnea index. Sleep 2009;32:150–7.
[19] Parrino L, Ferri R, Zucconi M, Fanfulla F. Commentary from the Italian association of sleep medicine on the AASM manual for the scoring of sleep and associated events: for debate and discussion. Sleep Med 2009;10:799–808.
[20] Moser D, Anderer P, Gruber G, Parapatics S, Loretz E, Boeck M, et al. Sleep classification according to AASM and Rechtschaffen & Kales: effects on sleep scoring parameters. Sleep 2009;32:139–49.
[21] Dyson R, Thornton C, Doré C. EEG electrode positions outside the hairline to monitor sleep in man. Sleep 1984;7:180–8.
[22] Terzano MG, Parrino L, Smerieri A, Chervin R, Chokroverty S, Guilleminault C, et al. Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep. Sleep Med 2002;3:187–99.